Weekly report (2017-9-4 ------ 2017-9-10)

Sept 10th, 2017, 19:12 pm

1. Progress

Table 1. Progress

Tasks	DUE DATE	TASKS IN PROGRESS
Dimensionality reduction	Sept 30th, 2017	Transform the t-SNE from single core into multicore, and test its time.

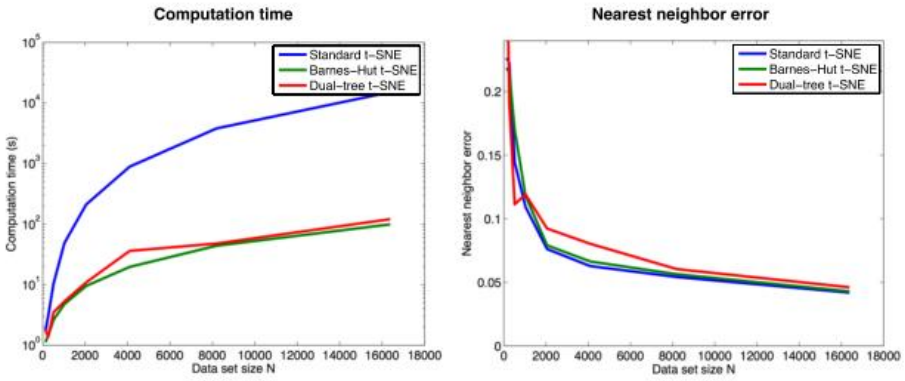
2. Research

2.1 paper reading

1. Visualizing high-dimensional data using t-sne

   This paper present t-SNE in 2008 and it can be compared with the next one "Accelerating t-SNE using Tree-Based Algorithms". The technique presented in 2008 was important for the high-dimensional data that lie on several different, but related, low dimensional manifolds. T-SNE can convert a high-dimensional data set into a matrix of pairwise similarities and visualize the resulting similarity data.

2. Accelerating t-SNE using Tree-Based Algorithms



In this paper, the author uses two tree based algorithms to accelerating t-SNE. The paper develops the Barnes-Hut of t-SNE and dual-tree algorithm of t-SNE, and make a comparison of these two algorithms. At last, the performance of Barnes-Hut variants of t-SNE better.

3. Fast approximate nearest neighbors with automatic algorithm configuration

$Pr.(\%)$	w_b	w_m	Algorithm	Dist. Error	Search Speedup	Memory Used	Build Time
60%	0	0	k-means, 16, 15	0.096	181.10	0.51	0.58
	0	1	k-means, 32, 10	0.058	180.9	0.37	0.56
	0.01	0	k-means, 16, 5	0.077	163.25	0.50	0.26
	0.01	1	kd-tree, 4	0.041	109.50	0.26	0.12
	1	0	kd-tree,1	0.044	56.87	0.07	0.03
	*	∞	kd-tree,1	0.044	56.87	0.07	0.03
90%	0	0	k-means, 128, 10	0.008	31.67	0.18	1.82
	0	1	k-means, 128, 15	0.007	30.53	0.18	2.32
	0.01	0	k-means, 32, 5	0.011	29.47	0.36	0.35
	1	0	k-means, 16, 1	0.016	21.59	0.48	0.10
	1	1	kd-tree,1	0.005	5.05	0.07	0.03
	*	∞	kd-tree,1	0.005	5.05	0.07	0.03

In the paper of the former one “Accelerating t-SNE using Tree-Based Algorithms”, it talks about an approach that automatically selects the best-performing tree-based algorithm for a particular data, so I find this paper. For many problems in computer vision, the most time consuming component consists of nearest neighbor matching in high-dimensional spaces. This paper also presents a new algorithm that applies priority search on hierarchical k-means trees, it has a good performance on many datasets. Besides, the author has found that multiple randomized k-d trees provide the best performance for other datasets.

3. For Prof. Chen

Happy Teachers’ Day!